

Assessing Achievement of Selected Objectives in Listening and Speaking

Executive Summary

MARCH, 1979

AE
ASS

Alberta

Minister's Advisory Committee
on Student Achievement



ASSESSING ACHIEVEMENT OF SELECTED OBJECTIVES
IN LISTENING AND SPEAKING

EXECUTIVE SUMMARY

ATA LIBRARY

A Study Conducted for
The Minister's Advisory Committee on Student Achievement
(MACOSA)

by a

RESEARCH TEAM OF THE UNIVERSITY OF CALGARY

Emma Platter - Faculty of Education
Department of Curriculum and Instruction
W. R. Unruh - Faculty of Education
Department of Educational Psychology
Laura Muir - Faculty of Education
Department of Curriculum and Instruction
K. D. Loose - Faculty of Continuing Education

through a contract with

ALBERTA EDUCATION

(Study administered by the Planning and Research Branch)

March, 1979



AE
ASS

CONDENSED SUMMARY

The Minister's Advisory Committee on Student Achievement (MACOSA) was established by Ministerial Order in October, 1976, to identify solid evidence of the level of achievement in Alberta schools (Alberta Education, 1977). The Committee's mandate included: (1) studying the elements of student achievement, including definition, purpose, principles, and assumptions underlying achievement; (2) consideration of ways and means to assess student achievement; (3) consideration of methods for maintaining or improving achievement; and (4) recommendations on these matters as based on findings.

In order to fulfill this mandate, MACOSA commissioned a number of studies in various achievement categories. On February 7, 1978, Alberta Education contracted The University of Calgary and specifically the principal investigators, E. Plattor, W. Unruh, and L. Muir, professors in the Department of Curriculum and Instruction and of Educational Psychology, to design and/or adapt tests for assessing the achievement of selected objectives in listening and speaking for Grades 3, 6, 9, and 12.

A Prospectus entitled Assessing Achievement of Selected Objectives in Listening and Speaking was submitted by the principal investigators to Alberta Education on January 10, 1978. This was followed by a request from Alberta Education for a formal Proposal for the study. A Proposal, also entitled Assessing Achievement of Selected Objectives in Listening and Speaking, was submitted on February 1, 1978. This Proposal was accepted by Alberta Education, and a contract between Alberta Education and The University of Calgary for a project entitled MACOSA Developing and Validating Achievement Tests in Speaking and Listening was finalized on February 7, 1978.

The following guidelines governed the study: (1) the tests would take a maximum of one hour of student time; (2) technological aids--tape recordings, etc.--would be used where feasible; (3) as much as possible, testing and observation would be conducted in natural settings; (4) statistical analyses would make use of traditional procedures for evaluation of items; (5) where appropriate, common stimuli would be used across at least two grades so that developmental trends could be observed; (6) when stimulus materials required the tape-recording of speaking activities, the voices of students and classroom teachers would be used in making the tapes; and, (7) where feasible, knowledges and attitudes would be tested with pencil-and-paper tests, while skills would be evaluated directly.

The test instruments in listening and speaking were designed for the following purposes:

1. To provide for program evaluation through the collection of baseline data in listening and speaking.
2. To provide for initial diagnostic assessment of individual students based on the achievement of the specified objectives. For this purpose, measures of individual achievement would be a secondary concern.
3. If possible, to provide teachers with models for their use in developing tests in listening/speaking based on specific objectives appropriate for their own students.

Thus, the purpose of this study was the development of criterion-referenced test instruments designed to assess Alberta (Canada) students' achievement of selected listening and speaking objectives at the Grades 3, 6, 9, and 12 levels. The principal investigators worked under the supervision of a Steering Committee appointed by the Minister's Advisory Committee on Student Achievement (MACOSA). The project was begun in February, 1978, and was completed in October, 1978.

Pilot, Field, and Final Test instruments in listening and in speaking at each of Grades 3, 6, 9, and 12 were constructed during the study, which consisted of the following phases: (1) selection, validation, and prioritization of measurable objectives in listening, speaking, and communication attitudes appropriate to each of Grades 3, 6, 9, and 12; (2) review of the literature in communication, listening, and speaking generally and measurement in listening, speaking, and communication attitudes specifically; (3) construction of Pilot Test instruments in listening, speaking, and communication attitudes for each of Grades 3, 6, 9, and 12; (4) administration of these Pilot Tests to samples of Alberta students at each grade level; (5) analysis of the data obtained from these Pilot Test administrations; (6) construction of Field Test instruments in listening, speaking, and communication attitudes at each grade level based on the findings of the Pilot Test data analyses; (7) administration of these Field Tests to samples of Alberta students at each grade level; (8) analysis of the data obtained from the Field Test administrations; (9) construction of Final Test instruments in listening and speaking for each

grade level based on the results of the Field Test data analyses. (These analyses indicated that the scales used to measure attitudes in this study did not perform this function or appear to be related to the types of activity used in the Listening and Speaking Tests, so this instrument was dropped.);and (10) development of a Technical Report of the study, as well as an Executive Summary, Condensed Report, and Abstract.

Knowledges, skills, and attitudes in listening and speaking were tested during the Pilot and Field Test phases. Since no usable tests or items in listening and speaking were located during the literature review, all items were constructed by the investigators for all of the Listening and Speaking Tests. Two commercial Communication Attitude Tests were administered in the Pilot and Field Test phases but were not included in the Final Tests.

The Final Listening Tests at each grade level consisted of a series of stimulus subtests designed to measure student skills in comprehending spontaneous talk, spontaneous prepared talk, and written language read aloud. Multiple choice items were used throughout. The tests were administered in an audio tape-recorded format. The items used were based on domains into which the selected objectives had been classified, as well as on the objectives themselves. Teacher's Manuals and Student Test Booklets were developed for use in administration at each grade level.

The Final Speaking Tests at each grade level consisted of two parts. Part I, Speaking (Written), consisted of stimulus subtests designed to measure student knowledges in both speaking and listening (communication competency). At the Grade 3 and 6 levels, Part I contained two

stimulus subtests, one designed to assess knowledge of selected techniques for the oral reading of prose and poetry, and the other to assess knowledge of organizational factors. At the Grade 9 level, Part 1 contained three stimulus subtests involving knowledge of selected techniques for the oral reading of prose and poetry as well as knowledge of organizational factors. At the Grade 12 level, this part contained two stimulus subtests involving knowledge of selected techniques for the oral reading of poetry only, as well as knowledge of organizational factors.

Part 2, Speaking (Oral), was designed to test student skills in oral reading (interpretation) and in spontaneous speech during an audience situation (a group of 4 - 6 of the students' peers). At the grade 3 and 6 levels, Part 2 of the Speaking Test involved the presentation by each student of a one minute talk based on a four-frame pictorial sequence, as well as the oral reading of the piece of prose analyzed in Part 1. At the Grade 3 level, Part 2 also contained an articulation subtest. At the Grade 9 and 12 level, Part 2 involved the presentation by each student of a prepared extemporaneous three minute talk based on either of the topics, "Hiking" or "Walking," as well as the oral reading of the poem analyzed in Part 1.

Multiple choice items were used in Part 1 of the Speaking Test, while Rating Scales were used to assess the tasks in Part 2. In addition, the spontaneous talks were tape recorded on audio tape. These were later analyzed again on an extended Rating Scale. The items used in both the multiple choice questions and the Rating Scales were based on domains into which the selected objectives had been classified as well as on the objectives themselves. Teacher's Manuals and

Student Test Booklets were developed for use in administration at each grade level.

Pilot testing was carried out in intact classes in Calgary and Okotoks, Alberta. Field testing was done in intact classes in Didsbury and Red Deer, Alberta.

The Pilot Test sample was the same for both the Listening and the Speaking Test administration. The sample consisted of 117 students in Grade 3, 124 students in Grade 6, 101 students in Grade 9, and 82 students in Grade 12.

The Field Test sample was also the same for both the Listening and Speaking Test administration. The sample consisted of 210 students in Grade 3, 235 students in Grade 6, 251 students in Grade 9, and 168 students in Grade 12.

Statistical analyses were performed on both the Pilot and Field Tests in order to determine: (1) if the options in the multiple choice items worked properly; (2) if the items were too difficult or too easy; (3) if the items discriminated among students; and (4) if the ratings and tests themselves were reliable.

In order to see to what extent the Pilot Tests met these criteria, item analyses, choice distributions, and inter-rater reliability coefficients were calculated for each of the tests. The results of these analyses are summarized below and on pages 7 to 9.

1. Items were evaluated for difficulty level. A cut-off of 0.30 to 0.80 in the Difficulty Index was arbitrarily established as a criterion. That is, items which were correctly answered by fewer than 30 percent or more than 80 percent of the students would be subject to revision or deletion.

2. Next, items were evaluated on the basis of their usefulness in discriminating between strong and weak students. A Discrimination Index level of 0.20 was applied.
3. Items in which options attracted fewer than five percent of student responses were also subject to revision.

Using this information as well as comments made by teachers, administrators, etc., the Research Team was able to revise the instruments for field testing. It was discovered that, indeed, there were a number of items that did not meet established criteria. Since these tests were longer than required, it was deemed possible to delete a number of weak items and still meet the requirements of a balanced test as well as of including at least two items to measure most of the objectives. Some items were retained even though they were easy because they measured high-priority objectives. The importance of objectives thus remained as the prime criterion in the item selection process for this test.

In order to test the reliability of the ratings made by the judges on the Speaking (Oral) Pilot Test items, a sample of the taped oral readings and speeches was selected and rated. This was done by selecting every fifth tape at each grade level. Two raters assessed each tape. Correlation coefficients ranged from 0.75 to 0.83. The average correlation was 0.78 for all four grades. It should be noted that items which required visual contact between rater and student were not included in this analysis. The results of these ratings were correlated to arrive at reliability figures.

In order to determine whether there was adequate validity in the ratings, choice distributions were also examined. There appeared to be

considerable spread in the ratings, indicating that raters did not simply assign the same score to students within any one scale. It was noted, however, that there were several items which raters frequently omitted. These items were revised by providing more adequate rating criteria and were included again in the Field Tests because it was felt that they measured important objectives.

Item analyses, descriptive statistics, and correlational analyses carried out for the Field Tests indicated that the Field Tests were valid in terms of measuring a variety of objectives; no reasonable way of re-designing the tests to make them shorter was apparent; and there appeared to be no reason that the attitude measure used here should be given again.

Since the domains were arrived at by an a-priori process, the Research Team felt that it would be useful to perform factor analyses on the Listening, Speaking, and Attitude Field Tests by grade level. Test statistics, item analysis data, and correlation data suggested that a male-female distinction in such an analysis would not be productive. The limited sample size, too, might lead to spurious results. Thus, a first-run factor structure based on the total population at each grade level was felt to be adequate and defensible. The results reported below are based on the use of the Varimax Rotation using an SPSS program.

Analysis of the Grade 3 Listening Test yielded six factors which accounted for 76 percent of the variance. These factors were labelled and compared to the original domain structure. A reasonably close

match was obtained, and it was concluded that the domains as originally conceived formed logical groupings of items for test interpretation.

Attempts were made to perform similar analyses of Grade 6, 9, and 12 Listening Test data. No reasonable factor structure could be found. Thus a large number of factors was required to account for substantial amounts of variance. It was concluded that the items on the test were heterogeneous--that is, they required a large number of abilities and skills which could not be grouped into meaningful categories. Since this was the case, no comparison of factor groupings to original domain structures could be made.

In the Speaking Tests, clear-cut factor structures emerged in Grades 3, 6, and 9. In each case six factors accounted for at least 90 percent of the variance. In Grade 12 a number of items, specifically those on the Written Test, did not load on any of the primary factors. This latter finding suggested that at this level the skills required to perform well on this test are of a heterogeneous nature and are largely dependent on the background and ability of individual students. The match between the obtained factors and the original domain structures was good. Again, it was concluded that the groupings of items in the original domains was a satisfactory base for interpretation.

The conclusion from all of the above analyses was that with minor revisions these Field Tests were now ready for administration as Final Tests to larger samples of Alberta students.

A number of recommendations based on the findings of this study were provided to MACOSA. These included the following:

1. The Speaking and Listening Final Tests should be given to larger samples of the Alberta school population. Further analyses should be performed on the results with a view of establishing acceptable performance criteria.

2. No possibility for shortening the Listening or Speaking Tests appears appropriate. Since these tests are based on stimulus sub-scales, and since these sub-scales contain a mixture of items/objectives measures, they must be given in intact form. Thus we recommend that the entire Listening Test and the entire Speaking Test for each grade level be given as presented. (However, each test in its entirety need not be administered in one session.)

3. Further statistical analyses should be done when data are derived from the Final Test samples. It is also suggested that analyses involving correlational techniques be used in order to devise sub-scales other than those based on a-priori domains.

4. While objectives for this study were carefully selected, further energy should be expended in continuous review and upgrading of objectives. A mechanism is needed for ensuring that such objectives become a part of the language arts/English curriculum. Such ongoing development would also require constant review and development of measurement techniques.

5. The use of the Personal Report of Communication Fear or the Personal Report of Communication Anxiety (McCrosky, 1977) is not recommended as part of the testing program in listening and speaking. These tests may well be useful as measures of communication anxiety, but their use should be limited to measuring only this aspect of student development. However, if MACOSA deems it necessary to provide

attitude tests for the listening/speaking areas of the language arts curriculum, a specific program for developing such tests should be undertaken. It is our opinion that tests of attitudes toward school, such as the ones currently being developed for MACOSA, might well be feasible.

6. Alberta Education should prepare a statement of those objectives it thinks are appropriate as part of the school program in listening and speaking at the elementary and secondary grade levels. Such a list should indicate which objectives are measured by the Final Tests prepared for this study. This statement should be made available to school systems throughout Alberta for use in future curriculum planning.

7. Pre-service and in-service training at all levels should be made available to teachers concerned with providing language arts/English instruction in elementary and secondary schools, in order that they can develop and expand the requisite skills for teaching and diagnosing listening and speaking knowledges and skills as well as for integrating listening and speaking into the total language arts/English curriculum. Further, pre-service and in-service training in the area of functional communication skills should be made available to all teachers, regardless of their curriculum speciality.

8. When the Final Listening and Speaking Tests are in general use in the province of Alberta, pre-service and in-service training should be provided to teachers in both the administration of the tests and the interpretation of test results. Further, training should be provided to teachers in the use of the test results in curriculum planning at the individual, class, school, and system levels.

9. The Listening Comprehension Tests developed in England by Wilkinson, Stratta, and Dudley (1974) would be a most useful addition to a Listening Test battery for Alberta schools. While these tests measure objectives not specifically dealt with in this study for reasons described in the Technical Report to MACOSA, they are based on objectives

considered highly relevant and valid by experts in the language arts field. If this test is to be used, Canadianization is essential.

ALBERTA SPEAKING TEST (MACOSA)

October, 1978

E. Plattor, W. Unruh,
L. Muir, K. Loose

The Speaking Pilot, Field, and Final Test instruments at each grade level consisted of a series of subtests designed to measure listening and speaking knowledges through multiple choice items and skill objectives through involvement in situations requiring spontaneous or spontaneous prepared talk. The three types of language measured in the Listening Test instruments were also measured in the Speaking Tests. At the Grade 3 level, articulation (ability to articulate standard English speech sounds) was also tested. Objectives were classified into domains, and stimulus subtests were constructed which included items or tasks representing these various domains. The Pilot, Field, and Final Speaking Test domains were: (1) Voice Production Factors; (2) Nonverbal Factors; (3) Linguistic Factors: Sounds; (4) Linguistic Factors: Words; (5) Linguistic Factors: Sentences; (6) Organizational Factors: Ideational Fluency; (7) Organizational Factors: Plan Analysis; and (8) Organizational Factors: Details.

At each grade level, the Final Speaking Tests consisted of two parts. Part 1, Speaking (Written), tested student knowledges in both listening and speaking (communication competency). Part 2, Speaking (Oral), tested student skills in speaking to a small group of their peers (about 4 - 6 students), as well as skills involved in oral reading (interpretation). At the Grade 3 level, Part 2 also contained an articulation test.

Part 1 of the Grade 3 and 6 Final Speaking Tests (Speaking Written) contained two stimulus subtests. One was designed to measure knowledge of how to read a poem for meaning and mood, while the other was designed to measure knowledge of how to read a piece of prose in terms of organizational factors. These subtests were planned to prepare students for the oral reading and the spontaneous speaking to be done in the second part of the Speaking Tests.

Part 1 of the Grade 9 Speaking Test contained three subtests, while the Grade 12 Speaking Test contained two subtests. At the Grade 9 level, the first subtest measured knowledge of how to read a poem and the second one measured knowledge of how to read a prose piece (interpretation). The third subtest measured knowledge of the organization of a speech. In this subtest, students' attention was directed to such organizational techniques as deciding on a main idea, selecting supporting details, etc. The subtests in Part 1 were designed to test knowledges in oral interpretation as well as to prepare students for the oral reading of the poem and the presentation of an extemporaneous speech during the second part of the Speaking Tests.

Thus the knowledges tested in Part 1 of the Speaking Tests at all grade levels were basically concerned with written language read aloud as well as with planned spontaneous language (in terms of organizational factors). Multiple choice items with three answer options were used to test these knowledges with a pencil-and-paper format.

Part 2 of the Grade 3 and 6 Speaking Tests (Speaking Oral) contained three subtests at the Grade 3 level and two subtests at the Grade 6 level. This part of the test was designed to assess skill in organizing and presenting a short (one minute) impromptu talk as well as to test skills involved in oral reading, and was prepared for administration to students in small groups (4 - 6 of their peers). At both grade levels, the first subtest involved the presentation of the impromptu talk. At each grade level, the talk was based on a four-frame pictorial stimulus. While the format was the same at both levels, the content of the frames was different at Grades 3 and 6.

During this test administration, each student in the group was in turn presented with the picture sequence and directed to examine it for about 30 seconds. (During this time, the other students in the group were recording their reactions to the previous speaker's talk on a special form provided for this purpose in the Student Test Booklet.) After he/she had examined the story frames, the student was asked to tell the other members of the group a story about the events depicted in the frames. While each student spoke, the teacher evaluated the talk using a Rating Scale constructed for this purpose. At the same

time, the teacher recorded the talk on audio tape, and later re-evaluated it using a more extensive Rating Scale. Thus this subtest was concerned with spontaneous talk.

The second subtest in Part 2 at the Grade 3 and 6 levels involved the reading of the piece of prose analyzed in Part 1. Thus this subtest was concerned with written language read aloud. At the Grade 3 level, an articulation subtest was included as well.

Part 2 of the Grade 9 and 12 Speaking Tests (Speaking Oral) contained two subtests at both levels. One subtest was designed to test skills in organizing and presenting an extemporaneous talk; and the other was designed to test skills involved in oral reading through the actual reading of the poem analyzed in Part 1. In the first subtest, given in small groups of their peers (about 4 - 6), all students in the group were given 15 minutes to prepare for a 3 minute speech. (In the Pilot Test, 45 minutes were given for this task. However, it was found that this provided sufficient time for students to write entire essays rather than talk outlines, and so the preparation time was reduced.) All students in the group prepared at the same time. Students were provided with a choice of two topics (Hiking or Walking); supporting materials in the form of quotations, statistics, etc.; and two worksheets, one of which provided assistance in selecting the purpose and audience for the talk, determining the language level, and narrowing the topic while the other provided an outline form for student use in developing notes for the talk. In the second subtest, five minutes were allowed for a re-examination of the poem previously analyzed in Part 1, and then each student in turn was directed to read the poem to the members of their group. (The groups for both subtests were the same.) Both the extemporaneous speech and the oral reading were evaluated by the teacher using special Rating Scales developed for the purpose. In addition, the teacher tape-recorded the extemporaneous speech and later re-evaluated it using a more extensive Rating Scale. Thus the two subtests in Part 2 were concerned with written language read aloud as well as with spontaneous prepared talk.

Teacher's Manuals for test administration at each grade level were constructed, as were Student Test Booklets. Special answer sheets were provided at the Grade 6, 9, and 12 levels. (Grade 3 students responded in the Test Booklets.) At all grade levels, the Field and Final Test administration times for Part 1 were 40 minutes. At the Grade 3 and 6 levels, the administration time for Part 2 was 65 minutes, and at the Grade 9 and 12 levels, the administration time was 80 minutes. (These times are based on 6 students at ten minutes per student for the presentation of the speech, as well as 15 minutes per group of students for preparation time.)

In the Grade 3 Field and Final Tests, 38 items and 23 objectives were tested through the use of multiple choice items and rating scales; at Grade 6, 28 objectives and 44 items; at Grade 9, 34 objectives and 53 items; and at Grade 12, 32 objectives and 50 items.

ALBERTA LISTENING TEST (MACOSA)

October, 1978

E. Plattor, W. Unruh,
L. Muir, K. Loose

The Listening Pilot, Field, and Final Test instruments at each grade level consisted of a series of subtests designed to measure the skills involved in comprehending three types of language: spontaneous talk, spontaneous prepared talk, and written language read orally. Objectives were classified into domains, and stimulus subtests were constructed which included items representing these various domains. The Final Listening Test domains were: (1) Voice Production Factors; (2) Linguistic Factors: Words; (3) Linguistic Factors: Sentences; (4) Organizational Factors: Literal Comprehension; (5) Organizational Factors: Interpretive Comprehension; and (6) Organizational Factors: Critical Comprehension. Two additional domains, tested in the Pilot Listening Tests were dropped from the Field and Final Tests in an effort to shorten the administration time. The objectives in these domains were combined with objectives in other domains or were dropped because they were found to be met by the Grade 3 level. (These domains were Environmental Sound Factors and Organizational Factors: Plan Analysis.)

Stimulus materials included all three language types, and stimulus subtests included items in the various domains. The multiple choice format was used throughout the Listening Tests. All tests were audio tape recorded and administration was done using these recordings. Teacher's Manuals for use in test administration at each grade level were constructed, as were Student Test Booklets. Special answer sheets were provided at the Grade 6, 9, and 12 levels. At the Grade 3 and 6 levels, the Field and Final Test administration times were 60 and 65 minutes respectively, while at the Grade 9 and 12 levels, the administrative time was 70 minutes. These times included preparation for the test administration. The actual tape times were 40 minutes at Grade 3; 45 minutes at Grade 6; and 60 minutes at Grade 9 and 12.

At the Grade 3 and 6 levels, the Field and Final Tests included 14 and 16 objectives and 35 and 40 items respectively. At the Grade 9 and 12 levels, the Tests included 20 objectives and 50 items each. Thus at least two items were used to test each objective at each grade level.

Date Due

THE ALBERTA TEACHERS' ASSOCIATION
11010 - 142 STREET
EDMONTON, — ALBERTA
T5N 2R1

371.26 Plattor, Emma et al - Assessing
A 846 achievement of selected objec-
1979 tives in listening and speaking:
c.2 executive summary

